Parallel Ab Initio Quantum Chemistry on Pentium® Pro Networks

Edward Seidl
Sandia National Laboratories

Introduction

- The need for speed
  - Quantum chemistry
    - **Traditionally scales as $O(N^4)$ - $O(N^2)$**
    - **Current workstations usable up to a dozen or so atoms**
    - **For hundreds of atoms need MPP machines**
  - MPP's
    - **Expensive**
    - **Not terribly robust**
    - **Shared with many users**

Introduction (cont'd.)

- Workstation clusters
  - Began as attempt to consume unused cycles
  - Fast PC's may lead to commodity supercomputers
  - Linux/GNU breaks dependence on costly software maintenance contracts
  - You control access

Specifics

- 14 ALR Quad Pro
  - 4 200MHz PPro with 256k cache
  - 256 Mb ECC FPM memory
  - DEC 21440 based fast ethernet card
  - Adaptec 2940W SCSI
  - 2Gb Seagate HD
- 2 Bay Networks fast ethernet switches
  - Linked to form virtual network
  - Can split cluster into two parts

Cost

Quantity	Description	Cost
1	3” Rack	$1,314
4	6” Rack	2,550
4	4-way switch	851
12	Switch box kit	74
14	ALR Quad6 w/ 4 200MHz PPro CPU	18,143
56	16x36 FPM SIMM	1,535
14	Adaptec 2940W SCSI adapter	258
14	SMC 9332 fast ethernet card	85
2	Seagate 4Gb hard drive	1,399
14	Seagate 2Gb hard drive	895
2	Bay Netwroks Lattis Switch 28115	12,572
4	MAG DX15 monitor	359
1	Nanao T2-17TS monitor	899
2	3Com 590 ethernet	119
	Total Cost for System	**$40,000**

Sandia National Laboratories
California Site

DAISy
Distributed Array of Inexpensive Systems

- 32-node dual Pentium Pro 200Mhz PC cluster
- UNIX compatible operating system (RedHat 4.0 w/2.0.27 Kernel)
- Network
  - **standard 10 Mb/s (10BASE-2) Ethernet (used for client node NFS mounts and any client node interactive work users find necessary)**
  - **switched 100Mbs/ (100BASE-TX) Fast Ethernet (used for user program message passing traffic)**
  - **Gb/s hypercube connected Myrinet (also used for program message passing traffic)**
- Motive
  - **investigate the viability of commodity PC technology to perform computation of scientific and engineering problems traditionally performed on "Supercomputers"**

Sandia National Laboratories
California Site

Megalon Software

- Linux
  - Freely available OS
  - SMP aware
  - Large user community
  - Source code available
- Edware 2.0.3
  - Originally hierarchy of tar files similar to SLS
  - Now use rpm
  - Stole much from Redhat
  - Move things around, PPro specific compiler options, fixes

Megalon Software (cont'd.)

- Tools
  - GCC, G++, G77, f2c, JDK 1.0.2
  - GDB
  - Emacs
  - CVS/RCS
- Access
  - Telnet/rsh/ftp
  - SSH
    - **Secure connections**
    - **Encryption**
    - **Scripts to replace rdist**
  - Java-based signup sheet
    - **Netscape IFC 1.0**
    - **No enforcement yet**

Performance

- Pentium®Pro is a decent performer
  - 720x720 matrix multiply compiled with gcc
    - **2 outer loops unrolled level 2 = 55 Mflops**
    - **Outer loop unrolled level 4 with strip mining = 91 Mflops**
    - **Above with B matrix transposed = 111 Mflops**
    - **Memory bandwidth can be a problem**

Performance

Speed of a 720x720 Matrix Multiply

N proc	2x2 transpose	4x1	4x1 transpose
1	55 Mflops	91 Mflops	111 Mflops
2	40 Mflops	90 Mflops	110 Mflops
4	20 Mflops	85 Mflops	105 Mflops

Mind Your Alignment

- Fast floating point requires that doubles are aligned on a two-word boundary
  - Bug in GCC - `malign-double` flag
  - Linux crt 1.o does not properly align stack
    - **Different execution times depending on number of environment variables**
  - Both problems fixed in Edware (fixed crt file in glibc-2.0.2 and libc-5.4.22)

Speed of a 300x300 Matrix Multiply on the Stack

	1x1	2x2 transpose	2x2 transpose pointers
Aligned	41 Mflops	50 Mflops	46 Mflops
Mis-aligned	18 Mflops	17 Mflops	15 Mflops

Communications

- Message Passing Interface MPI
  - MPICH - Argonne National Lab
    - **http://www.mcs.anl.gov/mpi/mpich**
  - LAM/MPI - Ohio Supercomputing Center
    - **http://www.osc.edu/lam.html**
  - I find LAM to be more convenient
  - Neither implementations is particularly fast.

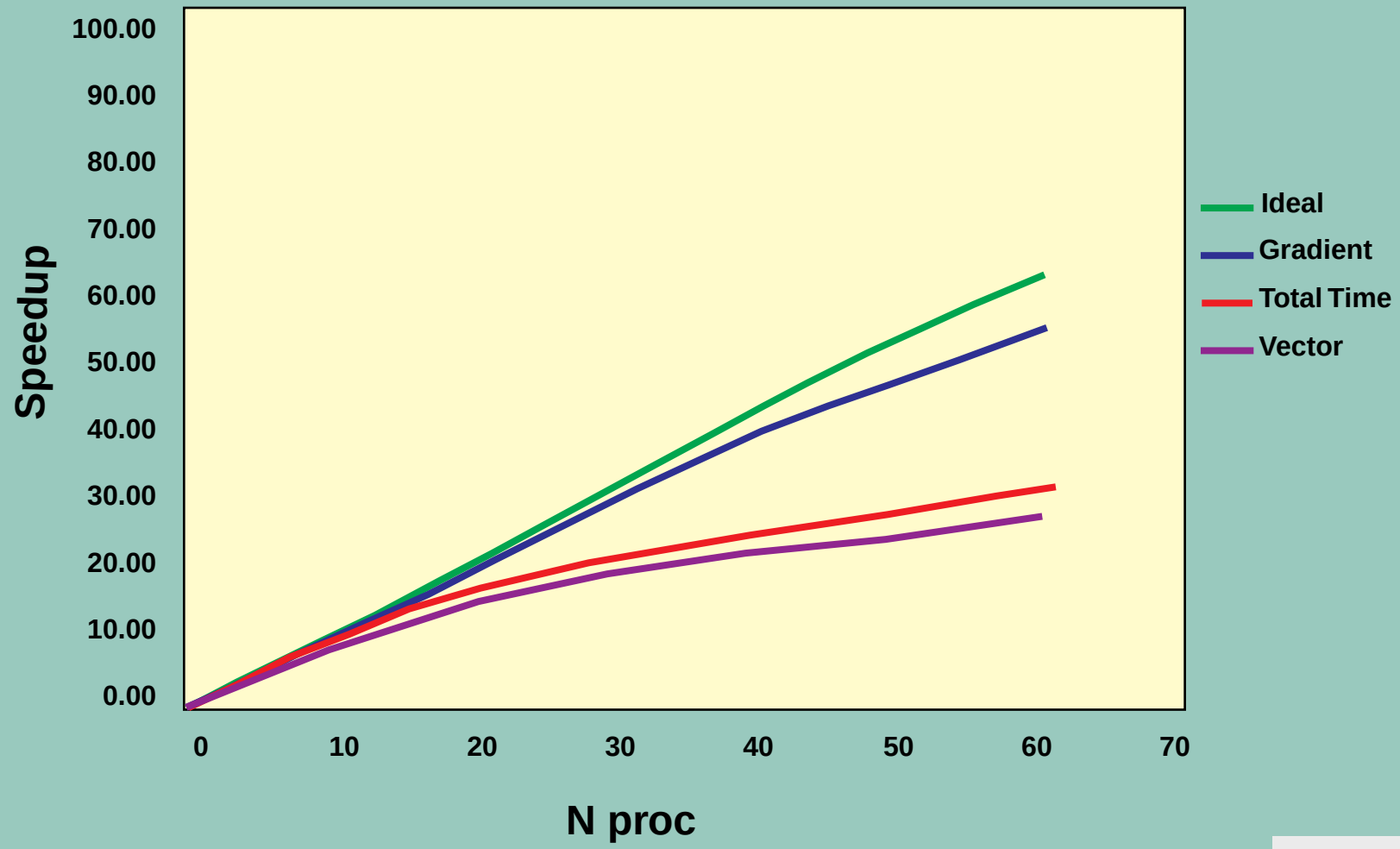
Performance of LAM/MPI

	Latency		Bandwidth	
N proc	normal (ms)	c2c (μs)	normal	c2c
1 per node				
2	2	200	33	22
4	2	200	33	22
8	2	210	32	22
2 per node				
2	2	74	45	95
4	3	300	9	20
8	3	300	7	20
16	3	300	9	20
4 per node				
4	4	180	23	69
8	5	550	6	12
10	5	550	5	12
32	5	550	5	12

Applications

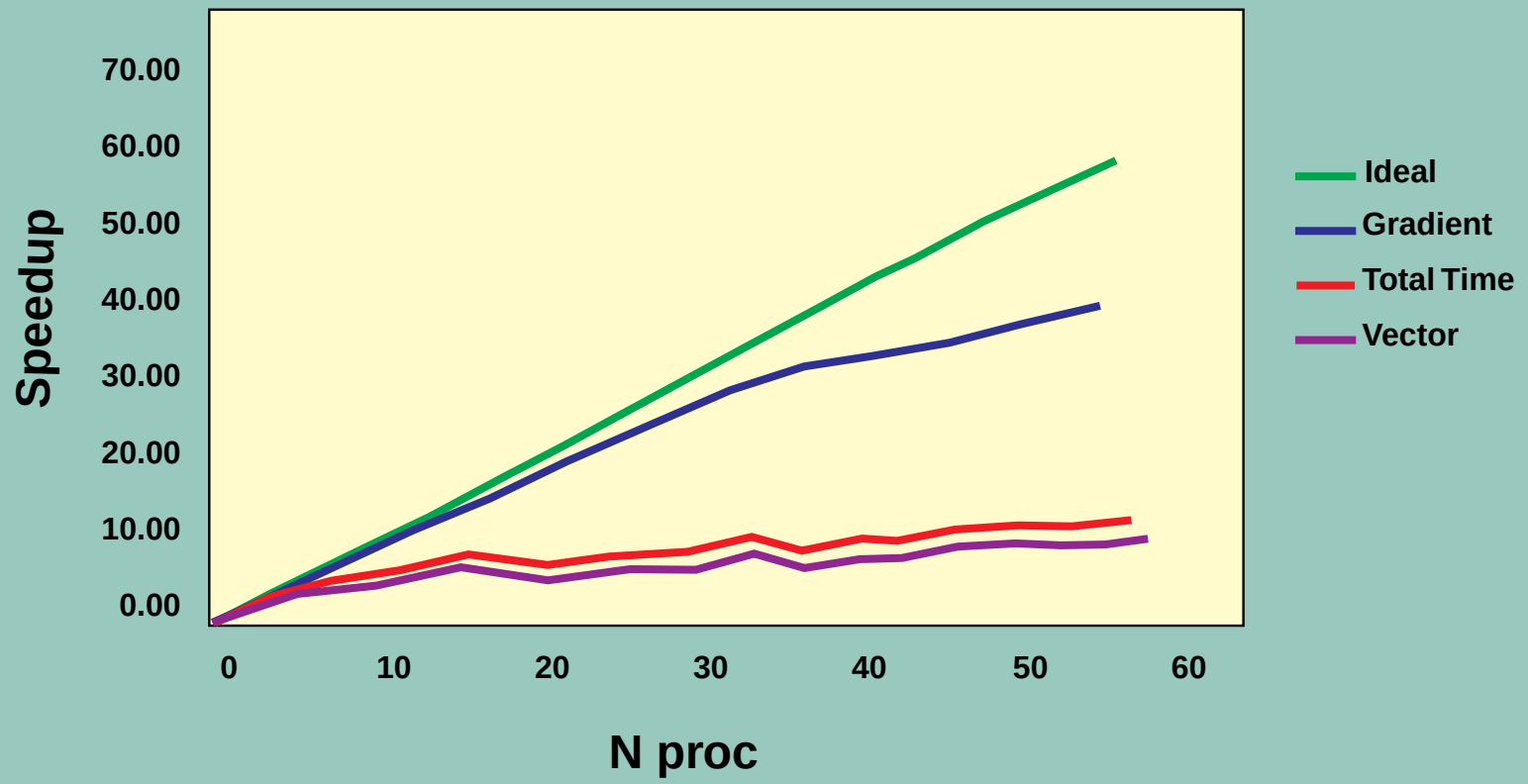
- MPQC - Massively Parallel Quantum Chemistry
  - Originally intended for MPP machines
  - Direct SCF and MP2
  - Energies and gradients
  - Finite displacement second derivatives

PETN CI Teraflops speedup



Sandia National Laboratories
California Site

PETN CI MPACTR speedup



Sandia National Laboratories
California Site

Multithreaded MPQC

- Use pthreads library
  - To get most up-to-date version installed glibc 2.0.1
  - Had to hack MT support into libg++
    - **Patches from H.J. Lu**
    - **A few things missing**
  - Works fairly well

Conclusions

- Pentium Pro clusters give you plenty of bang for your buck
- Dual systems may be a better way to go depending on the application
- Direct access to the network is a must